Creating the area-weighted average PM2.5, accounting for different buffer sizes of PM monitors

1) Merge monitor information (area coverage) from AQS database with monitors in PM2.5 dataset – by state, county, tract, site ID, and POC (unique at this level)
2) Now have PM2.5 dataset with monitor area information. Create buffers based on area coverage information.
3) For each day, get the intersection between these monitors to see how many of them overlap one another. Rename n.overlaps as buffer_overlaps, origins as buffer_origins. We do this because we're doing another intersection that would overwrite these variables.
4) Now, we want to aggregate (get area-weighted average) PM2.5 at the census tract level. To do this, for each day, do an intersection between the monitor intersections (3) and census tract shapefile.
    a. Create a monitor_count variable (count of monitors that a census tract intersects with in a given day).
    b. For census tracts covered by only 1 monitor in a day (monitor_count=1), PM2.5 is as is.
    c. For those covered by 2 or more monitors in a day (monitor_count>1):
        i. If buffers don't overlap (buffer_overlaps=1 or =2). Get area covered by each monitor, sum to get the aggregate.
            - Weight of each monitor value (per day): Weight = int_area/agg_area
            - $wgtd\ PM2.5 = PM_1 \left(\frac{int\ area_1}{agg\ area}\right) + PM_2 \left(\frac{int\ area_2}{agg\ area}\right) + \ldots + PM_n \left(\frac{int\ area_n}{agg\ area}\right)$
        ii. If buffers overlap (buffer_overlaps>2),
            - Weight of each monitor value (per day): Weight = int_area/agg_area
            - If the overlap is not fully overlapped (i.e when monitor buffers do not have same geometry/size/area and long/lat):
                a. Intersected shape has 1 or 2 buffer origins (implies only 1 of the multiple monitors covers the specific intersection), PM2.5 is as is
                b. For <u>parts of intersected shape that has more than 2 buffer origins</u> (implies many buffers cover that intersection), need to get the simple average PM2.5 value of those monitors
                    i. R assigns only 1 PM value for that shape, so we need to find the other monitor(/s) that covers the shape and get their mean
                    ii. We start by unnesting the buffer_origins by date and tract. We do this for monitor_count>1 (include buffer_overlap==2 since they usually contain the other PM2.5 value we need to get the average).
                    iii. Aggregate the unnested subset by tract, date, and buff_origin. By doing this, we get the averaged PM2.5 values for areas with multiple monitor overlaps in a day
                c. $wgtd\ PM2.5 = PM_1 \left(\frac{int\ area_1}{agg\ area}\right) + Mean(PM_2 + PM_3 + \cdots + PM_i) \left(\frac{int\ area_2}{agg\ area}\right) + \ldots + PM_n \left(\frac{int\ area_n}{agg\ area}\right)$
5) From these, we should have area-weighted PM2.5 for each date and tract